DOCUMENT RESUME

ED 223 144 HE 015 572

AUTHOR Abrami, Philip C.; Mizener, Deborah A.
TITLE Student/Instructor Attitude Similarity, Course Ratings and Student Achievement.
PUB DATE Aug 82
NOTE 25p.; Paper presented at the Annual Meeting of the American Psychological Association (Washington, DC, August 23-27, 1982).
AVAILABLE FROM Department of Education, Concordia University, 1455 de Maisonneuve Boulevard. W., Montreal, Quebec, Canada H3G 1M8.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Course Evaluation; Foreign Countries; Grades (Scholastic); Postsecondary Education; *Student Attitudes; *Student Evaluation of Teacher Performance; Student Teacher Relationship; *Teacher Attitudes; Teacher Effectiveness; *Teacher Influence
IDENTIFIERS Canada; *Quebec

ABSTRACT

The relationship between course ratings, student achievement, and the perceived attitude similarity of students and their instructors were explored with 645 students enrolled in 41 courses taught at English-language institutions in Montreal, Quebec. These institutions provide instruction intermediate between secondary school and the university. Two groups were evaluated. Three hundred and forty-nine students enrolled in 22 classes rated instructor effectiveness on a 28-item rating form and then rated their attitudes plus perceived instructor attitudes on an attitude scale adopted from Byrne. Another 296 students enrolled in 19 classes rated instructor effectiveness and their attitudes plus perceived instructor attitudes on an attitude scale adopted from Hofman and Kremer. Official end-of-term grades were also collected. A modest correlation between students' ratings and perceived attitude similarity was found, as was a modest relationship between similarity and course grades. These relationships were greatly reduced when the substantial influence of teachers on ratings and achievement was removed. The findings fail to support the claim that attitude similarity is a source of bias in ratings with any useful, practical effect. For neither of the attitude scales were student attitudes substantially related to student ratings of achievement independent of teacher effects. However, for the Hofman and Kremer scale, perceived instructor attitudes substantially affected ratings but not achievement. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223144

July, 1982

STUDENT/INSTRUCTOR ATTITUDE SIMILARITY, COURSE RATINGS AND STUDENT ACHIEVEMENT

Philip C. Abrami & Deborah A. Mizener

Concordia University, Montreal, Canada McGill University, Montreal, Canada

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Philip C. Abrami

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper presented at the annual meeting of the American Psychological Association, Washington, D.C., August 1982.

This work was partially sponsored by a grant to the first author from Formation de chercheurs et action concertée, Ministère de l'Éducation, Gouvernement du Québec.

Requests for reprints should be addressed to:
Philip C. Abrami, Department of Education, Concordia University
1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada H3G 1M8

HE 015 572

ABSTRACT

We explored the relationships between course ratings, student achievement and the perceived attitude similarity of students and their instructors. 349 students enrolled in 22 classes first rated instructor effectiveness on a 28-item rating form and then rated own attitudes plus instructor attitudes on an attitude scale adopted from Byrne. Another 296 students enrolled in 19 classes rated instructor effectiveness and own attitudes plus instructor attitudes on an attitude scale adopted from Hofman and Kremer. Official end-of-term grades were also collected. Consistent with prior research, we found modest-sized correlations between student ratings and perceived attitude similarity.There was also a modest relationship between similarity and course grades. These relationships were greatly reduced when the substantial influence of teachers on ratings and achievement were removed. The findings fail to support the claim that attitude similarity is a souce of bias in ratings with any useful, practical effect.

STUDENT/INSTRUCTOR ATTITUDE SIMILARITY, COURSE RATINGS AND STUDENT ACHIEVEMENT

Despite their widespread use in higher education and a voluminous research literature, a myriad of controversies continue to surround student ratings of instruction, whether they are used for promotion and tenure decisions or for course improvement. Many of these controversies concern the validity of student ratings. In the present paper we use a popular, but not universally accepted definition of validity. We consider student ratings to be valid to the extent that ratings predict instructor impacts on teacher effectiveness criteria, primarily student achievement but also motivation, interest, and attitude toward the subject matter. We consider student ratings to lack validity to the extent that ratings are influenced by factors shown to be theoretically or empirically irrelevant or biasing--factors which can be shown through logical argument or experimental evidence to affect student ratings without similarly affecting teacher effectiveness criteria.

As the title of our paper suggests we are specifically concerned with the relationships between student/instructor attitude similarity, course ratings and student achievement. This concern grew out of Feldman's (1977) exhaustive review of the literature on the consistency and variability among students in rating their professors. In that review, Feldman concluded that the assumed or perceived similarity of students and their instructors in attitudes, personality traits, etc. is related to course ratings. Furthermore, the evidence favoring a relationship between

student/instructor similarity and ratings appeared somewhat stronger than the evidence for a relationship between individual student characteristics and ratings.

Unfortunately, conclusions about the validity of ratings based on prior research on student characteristics are difficult to draw because of methodological flaws in prior studies. We (Abrami, Perry, & Leventhal, 1982; Abrami & Mizener, in press) have described these problems along with Marsh (in press) and Feldman (1977) as: a) using correlational evidence to suggest causation, b) failing to distinguish practical significance from statistical significance, c) inattention to the control of error rate per experiment in large multivariate investigations, and d) selection of the inappropriate unit of analysis, confounding variance attributable to students with variance attributable to instructors.

In two prior investigations we attempted to avoid these problems. In Abrami et al. (1982) we investigated the relationship between student personality characteristics, teacher ratings, and student achievement. In addition, we investigated the perceived similarity of instructor and student personalities by analysing the components of a similiarity score--each student's own score and the instructor's score as perceived by each student. We found no meaningful and consistent relationship between ratings and student personality characteristics independent of teacher effects. However, we did find that the personality characteristics of instructors, as perceived by students, were related to ratings, particularly so for perceived need to achieve, a factor we considered particularly relevant to an instructional setting.

In Abrami and Mizener (in press) we investigated the relationship between student/instructor attitude similarity and course ratings, but not student achievement. University students completed a course evaluation and a broadband attitude scale, once describing their own attitudes and once describing instructor attitudes. There was a modest correlation between attitude similarity and ratings which was substantially reduced when teacher effects on ratings were removed. The similarity of student and instructor attitudes on topics relevant to higher education was significantly related to course ratings, while the similarity of attitudes on topics less relevant to higher education was not. Of the components of a similarity score, only the perceived attitudes of the instructor accounted for a significant amount of rating variance independent of other effects. We concluded that the unique variance explained by student/instructor attitude similarity was of minimum importance when evaluations are used for summative purposes to make gross distinctions among instructors.

Our findings and conclusions stand in contrast to those of other researchers, most notably Hofman and Kremer (1980). In Study 1, Hofman and Kremer regressed global course ratings on a) student own attitudes toward higher education and teaching and b) the actual attitude similarity between student self-assessment and instructor self-assessment. Student attitudes and student/instructor attitude similarity accounted for 17.7% of the variance in ratings. In Study 2, Hofman and Kremer regressed global ratings on a) student own attitudes toward college teaching, b) class attitudes as perceived by the instructor, c) instructor own

attitudes, d) instructor attitudes perceived by each student, e) and four similarity indices (i.e., a with c, a with b, c with d, and b with d). The four sets of attitude items a-d and the four sets of similarity scores together accounted for almost 50% of the variance in ratings. Perceived instructor attitudes accounted for 70% of the criterion variance explained by the combined predictors and was the most efficient subset. Hofman and Kremer concluded:

> Course ratings are a function of both students' own attitudes and, even more, of interactions between students' and instructors' attitudes. Hence, attitudes are to be added to the list of personal variables that compete with situational ones in explaining evaluation variance. Surely some instructors are in fact more effective than others, but student ratings are limited in their ability to reflect those differences (and those differences alone). In any case, the validity of course evaluation as a criterion of instructional standards needs to be modified and investigated further. (1980, p. 616).

Unfortunately, Hofman & Kremer used pooled students as the units of analysis in a correlational design where teacher effects on ratings were not separated from student and other effects. Thus, inferences about whether students are limited in their ability to perceive differences in teaching ability cannot be readily drawn. Therefore, the purpose of the current study was to continue the line of inquiry begun in our prior studies and to resolve inconsistencies between our findings and conclusions and those of Hofman and Kremer (1980). More specifically, we

wished to replicate the findings of Abrami and Mizener (in press) with a different, more heterogeneous sample of students, courses, and instructors. We also wanted to include a measure of student achievement to determine the relationship between attitude similarity and course performance. Finally, we wanted to replicate and re-interpret the findings of Hofman and Kremer by administering a modified version of their attitude scales.

Method

Subjects

The subjects were 645 students enrolled in 41 courses taught at English-language CEGEPs in Montreal, Quebec. CEGEPs are institutions providing instruction intermediate between secondary school and university. Students who continue on to university need to complete only three years of full time study to earn a bachelor's degree.

Instruments

We used the Abrami and Mizener (in press) revised version of Byrne's (1971) Survey of Attitudes as one instrument to assess student own attitudes and student perceptions of instructor attitudes. The scale consists of 24 items, each rated on a six-point scale from strongly agree (like, etc.) to strongly disagree (dislike, etc.). Each item was selected for the scale because it had been judged high in topic importance and low in ambiguity. The range of topics covered by the Byrne scale is quite broad and includes politics, health, religion, as well as education. From these 24 items, Abrami and Mizener selected six which were relevant to higher education (e.g., grades, professor and student needs) and six which were irrelevant (e.g., birth control, inflation). These groups of

items did not differ significantly ($p > .05$) in either topic importance or item ambiguity.

We employed 18 items from Hofman and Kremer to measure student attitudes toward higher education and college teaching. Fourteen items were slightly modified from their 16-item "Attitudes Toward Higher Education." Two items from their scale were dropped because they failed to load highly on any of the four factors associated with their scale. Four items were selected from their 8-item "Attitudes Toward College Teaching." Three items from their scale were dropped because they were redundant with items from their higher education scale. The two remaining items, which covered the same topic but were oppositely worded, were combined into a single item.

In contrast to Hofman and Kremer where items responses could range from low agreement to high agreement, we chose a five-point response scale anchored by strongly disagree and strongly agree, to allow students to express uncertain as well as dissenting views.

The teacher rating form (TRF) consisted of 28 items each rated on a 5-point scale where a high score was a positive rating. Three questions represented global items of teaching effectiveness. The remaining items were adapted from Hildebrand and Wilson (1970) and asked about specific aspects of teaching.

Factor analysis of the 25 items by Hildebrand and Wilson (1970) revealed a five-factor solution identifying components of effective teaching as perceived by students. Factor 1 was labelled "Analytic/ Synthetic Approach" and relates to instructor scholarship, including

breadth, conceptual understanding and analytic ability. Factor 2 was labelled "Organization/Clarity" and relates to an instructor's presentation skill including clarity and preparation. "Instructor-Group Interaction" was the label for Factor 3. It relates to rapport with the class as a whole. Factor 4 was labelled "Instructor-Individual Student Interaction" and relates to the rapport between the individual student and the instructor. Factor 5 was labelled "Dynamism/Enthusiasm." It relates to the flair and excitement evidenced in an instructor's teaching.

Procedure

We randomly divided the classes so that 349 students enrolled in 22 classes completed the Hofman and Kremer attitude scale and 296 students enrolled in 19 classes completed the Byrne attitude scale. Questionnaires were administered from midsemester to near semester's end during single class periods during the Spring, 1981. First, students rated their instructor on the TRF. Second, students completed an attitude scale, rating their own attitudes. Finally, each student completed an identical attitude survey giving their impressions of the instructor teaching the course. We also collected official course grades at the end of term, where grades were distributed on a 100-point scale.

Data Analysis

For each attitude item, an individual similarity score was the difference between student self-ratings and student perceptions of teacher attitudes. A total similarity score (TOTSIM) was computed by adding the absolute differences of the individual similarity scores on all attitude items from either the Byrne scale or the Hofman and Kremer scale. We did not

compute a similarity score based on the D statistic (Osgood, Suci, & Tannenbaum, 1957) because Abrami and Mizener (in press) found no differences between TOTSIM and D statistic analyses.

Seven measures of each student's general impression of his/her instructor's effectiveness were computed: the average response to the first three global TRF items (TRFA), the average response to the remaining 25 specific TRF items (TRFB); and the average responses to the five items loading highest on each of the dimensions of teaching effectiveness identified by Hildebrand and Wilson (1970).

Multiple regression analyses were performed separately for each attitude scale using subprogram REGRESSION from the SPSS package of computer programs (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). The first analyses were the regressions of TRF scores on total similarity scores. The second analyses, designed to determine the relationships between similarity and TRF scores independent of teacher effects, were stepwise regressions where teacher effects were entered first followed by total similarity. In these and subsequent analyses, dummy variable coding was used to create vectors representing the different classes so that students within a class would be uniquely associated with that class. By specifying the order of entry a priori, we insured that the rating variance attributed to student/teacher similarity on the second step of the regression was independent of between class teacher effects. In contrast, the rating variance attributed to teachers included not only the unique variance attributable to instructor differences across classes but also variance shared between teacher effects and similarity effects.

Next and only for the Byrne data, we regressed TRF scores on the similarity scores for the relevant items independent of teacher effects. We repeated these analyses for the similarity scores based on the irrelevant items.

Additional analyses were designed to assess the predictive ability of the component parts (own attitudes and perceived attitudes) of a similarity score. Stepwise regressions were performed where teacher effects were entered on the first step and each component score alternately entered on the second or third step. Unlike a total similarity score, it did not seem conceptually reasonable to compute a total score for either the own attitude or perceived attitude scores. For these later analyses, we entered each component as sets of individual scores corresponding to the number of attitude items, and did not use the TOTSIM as a similarity index. Instead, we employed the set of individual similarity scores as predictors to reduce the possibility of bias due to differences in the number of predictors used. In addition, individual similarity scores, because they were simple difference scores were perfectly collinear with the component scores. TOTSIM scores, because they were summed absolute difference scores, were not perfectly collinear with the component scores. For each stepwise regression, we used the mean square of the residuals on the last step as the error term for that analysis.

We repeated the stepwise analyses using student course grade as the criterion measure, entering the class vectors first in each analysis. Class-to-class differences in student achievement due to teaching quality as well as grading idiosyncrasies were, therefore, removed before similarity or component scores were analysed.

Results

Byrne Attitude Scale

The assigned teacher ratings were similar to those reported by Abrami and Mizener (in press)--somewhat more positive than those expected for an average instructor, but with a narrower than expected spread of scores. The mean rating for TRFA was 3.45; the SD was 0.69. The mean TRFB rating was 3.58; the SD 0.48. TOTSIM scores for the Byrne scale could range from 0, total agreement between student and instructor on all 24 attitude items, to 120, maximum disagreement with opposite and extreme views held. There was some disagreement between students and instructors but not a great deal. The mean TOTSIM score was 24.56; the SD was 9.98.

All simple correlations between TOTSIM and the seven TRF scores were significant ($p < .05$) and moderately sized (the average r was -.20): ratings improved as student and instructor attitudes became more closely matched. However, the variance in TRF scores explained by TOTSIM was reduced when teacher effects on ratings were removed (see Table 1).

Insert Table 1 about here

When the teacher vectors were entered first in the stepwise regressions, the variance in TRF scores explained by TOTSIM became even smaller, with .017 the average change in R^2, which is an average decrease of 60.3% in explained TRF variance. In contrast, the TRF variance (R^2) explained by the teacher vectors entered on the first step were quite large: .249

for TRFA, .312 for TRFB, .185 for Analytic/Synthetic, .303 for Organization/Clarity, .374 for Instructor-Group Interaction, .311 for Instructor-Individual Student Interaction, and .389 for Dynamism/Enthusiasm. All of these findings are similar in size and direction to those reported by Abrami and Mizener.

The next set of analyses compared the TRF variance explained by the similarity scores formed from the six relevant items to the similarity scores formed from the six irrelevant items after removing teacher effects. Similarity scores formed from the relevant items explained a significant ($p < .05$) but small amount of variance (average $\Delta R^2 = .015$) in four of seven TRF scores. For no TRF score was this significantly more than the variance explained by the similarity scores based on the irrelevant items.

The final sets of regression analyses for the TRF criteria compared individual similarity scores to the sets of component scores. The set of 24 individual similarity scores did not explain a significant ($p > .05$) amount of TRF variance. Neither set of component scores--student own attitudes or perceived instructor attitudes--added significantly to the prediction of ratings after teacher effects on ratings were removed.

The results for the relevant items and perceived teacher attitudes were different than those reported by Abrami and Mizener. We wondered whether the involvement of students and instructors with the issues raised by the attitude scale were different in the two studies and might explain the different results. While we could not compare the two samples for perceived importance of the issues raised, we could compare the samples for extremity of attitudes held to determine whether the present

sample held positions which were more neutral. Only four of 24 comparisons of student own attitudes were significant in the predicted direction, one-tailed $t(633) > 1.797$, $p < .05$. The mean difference for these four items was 0.262 scale points. Only six of 24 comparisons of perceived instructor attitudes were significant in the predicted direction, one-tailed $t(633) > 1.653$, $p < .05$. The mean difference for these six items was 0.192 scale points. It does not appear that differences across the studies in student attitudes or perceived teacher attitudes can account for the inconsistent findings.

The analyses of the achievement data were based on the course grades of 221 students enrolled in seventeen courses. The mean final grade was 74.60; the SD was 12.38. We consider this a fairly typical grading distribution. The simple correlation between grades and TOTSIM was $r = -.248$. Students whose attitudes matched their instructor's received higher grades than dissimilar students. However, when we controlled for class-to-class differences in teachers, the size of this relationship changed. TOTSIM accounted for a significant, but smaller, amount of achievement variance, $F(1,203) = 8.932$, $p < .01$, $\Delta R^2 = .035$. In contrast, the achievement variance explained by the teacher vectors entered on the first step was large ($R^2 = .180$). The similarity score based only on the relevant items also accounted for a significant amount of achievement variance, $F(1,211) = 9.002$, $p < .01$ but not more than explained by the similarity score based on the irrelevant items ($p > .05$). The set of 24 individual similarity scores did not explain a significant ($p > .05$) amount of achievement variance. Of the component scores neither student own attitudes nor perceived

teacher attitudes added significantly to the prediction of final grades.

Hofman and Kremer Attitude Scale

The mean teacher ratings were 3.32 for TRFA and 3.58 for TRFB; the SD's were 0.72 and 0.45, respectively. TOTSIM scores for the Hofman and Kremer scale could range from 0 to 72. The mean TOTSIM score was 14.01; the SD was 5.77. The descriptive TOTSIM statistics for the Hofman and Kremer scale resemble those for the Byrne scale, once differences in the total number of items and number of response alternatives per item are considered.

All simple correlations between TOTSIM and the seven TRF scores were significant ($p > .05$) and moderately sized with the average $r = -.26$. Teacher ratings were higher when students perceived that their attitudes were similar to their instructor's attitudes.

TRF variance explained by TOTSIM was reduced when teacher effects on ratings were removed (see Table 2). When the teacher vectors were entered

Insert Table 2 about here

on the first step of the regressions, the variance in TRF scores explained by TOTSIM averaged .034 (ΔR^2), which is an average decrease of 52.0% in explained TRF variance. On the other hand, the variance in ratings explained by the teacher vectors on the first step were large: .272 for TRFA, .276 for TRFB, .186 for Analytic/Synthetic, .353 for Organization/Clarity, .387 for Instructor-Group Interaction, .311 for Instructor-Individual Student Interaction, and .322 for

teacher attitudes added significantly to the prediction of final grades.

Hofman and Kremer Attitude Scale

The mean teacher ratings were 3.32 for TRFA and 3.58 for TRFB; the SD's were 0.72 and 0.45, respectively. TOTSIM scores for the Hofman and Kremer scale could range from 0 to 72. The mean TOTSIM score was 14.01; the SD was 5.77. The descriptive TOTSIM statistics for the Hofman and Kremer scale resemble those for the Byrne scale, once differences in the total number of items and number of response alternatives per item are considered.

All simple correlations between TOTSIM and the seven TRF scores were significant ($p > .05$) and moderately sized with the average $r = -.26$. Teacher ratings were higher when students perceived that their attitudes were similar to their instructor's attitudes.

TRF variance explained by TOTSIM was reduced when teacher effects on ratings were removed (see Table 2). When the teacher vectors were entered

Insert Table 2 about here

on the first step of the regressions, the variance in TRF scores explained by TOTSIM averaged .034 (ΔR^2), which is an average decrease of 52.0% in explained TRF variance. On the other hand, the variance in ratings explained by the teacher vectors on the first step were large: .272 for TRFA, .276 for TRFB, .186 for Analytic/Synthetic, .353 for Organization/Clarity, .387 for Instructor-Group Interaction, .311 for Instructor-Individual Student Interaction, and .322 for

Dynamism/Enthusiasm.

For only one TRF criterion score, Instructor-Individual Student Interaction, was there a significant increment in explained variance, $F(18,309) = 2.006$, $p < .01$, when the set of 18 individual similarity scores entered the equation after the teacher vectors. Of the component scores, when student attitudes entered on the third step (following the teacher vectors and perceived teacher attitudes) only two of seven effects were significant. (See Table 3).

Insert Table 3 about here

When perceived instructor attitudes entered on the last step, the results were different. Five of seven effects were significant and moderately sized (average change in $R^2 = .091$). For each of the five criteria, the variance explained by instructor attitudes was significantly greater than that explained by student attitudes, $F(1,291) > 13.00$, $p < .001$.

When Hofman and Kremer (1980) regressed ratings onto the set of perceived instructor attitudes toward college teaching, they accounted for 32.4% of the variance. When we entered the set of perceived instructor attitudes on the first step, the explained TRF variance (R^2) was much larger than the last step results, but less than that reported by Hofman and Kremer: .069 for TRFA, .199 for TRFB, .118 for Analytic/Synthetic, .107 for Organization/Clarity, .285 for Instructor-Group Interaction, .241 for Instructor-Individual Student Interaction, and .142 for Dynamism/Enthusiasm.

The analyses of the achievement data were based on the course grades of 254 students enrolled in 19 classes. The mean final grade was 75.39; the SD was 12.90. The simple correlation between grades and TOTSIM was -.145. Somewhat higher grades were received by students whose attitudes matched those of their instructors. This small relationship became insignificant ($p > .05$) when we controlled for class-to-class differences in teachers. In contrast, the achievement variance explained by the teacher vectors was large ($R^2 = .216$). Neither the set of student attitudes nor instructor attitudes accounted for a significant increment in achievement variance after the inclusion of teacher effects.

Discussion

Consistent with prior research, we found significant, although modest, correlations between student ratings and the perceived attitude similarity of students and their instructors. There was also a modest relationship between attitude similarity and course grades. Thus, students whose attitudes more closely resemble the attitudes of their instructors earn somewhat higher grades and assign somewhat higher teacher ratings than students with dissimilar attitudes.

Disregarding how these correlations came about, but noting their size and direction, leads one to conclude that the validity of student ratings is not substantially affected by student/instructor attitude similarity. Ratings and achievement were affected similarly by perceived attitude similarity.

Now let us consider the factors which affect ratings and achievement. In our research we did so by first removing teacher effects from ratings

and course grades. We considered the effects of teachers to be a relevant source of influence on ratings, particularly when class-to-class differences in instruction also affect course grades. In both samples of classes we found moderately large teaching effects on ratings and smaller, but also moderately large, teaching effects on achievement. As in prior research (e.g., Abrami, et al., 1982; Leventhal, Perry, & Abrami, 1977) ratings appear more sensitive to differences in instruction than student achievement. However, one important caution for this field research: what we describe as teacher effects include the unique effects of teaching ability on both criteria and also variance shared with other effects including attitude similarity.

After removing the effects of teachers, we determined the unique influence of attitude similarity on ratings and achievement. If we found that attitude similarity explained substantially more variance in ratings than achievement or vice versa, we would be in a better position to claim that attitude similarity was a source of bias in ratings. However, the unique effect of attitude similarity on ratings was small and not much greater than the effect on achievement. Even if one argues that any unique effect of attitude similarity on ratings is biasing regardless of its effect on achievement, the findings do not suggest that attitude similarity is a source of bias in ratings with any useful, practical implications.

What of the components of a similarity score? For neither the Byrne attitude scale nor the Hofman and Kremer attitude scale were student attitudes substantially related to student ratings or achievement independent of teacher effects. However, for the Hofman and Kremer scale,

perceived instructor attitudes substantially affected ratings but not achievement.

Whether perceived instructor attitudes are a source of bias in ratings is not clear. In this and prior studies, it is impossible to determine whether student differences in their perceptions of instructor attitudes affect student ratings or if the direction of the causal relationship is reversed (or even due to some other factor). It is conceivable to imagine students who form their judgements of an instructor's teaching ability and then generalize this view to include the instructor's attitudes and values. We share with Hofman and Kremer (1980) the hope that future studies attempt to investigate these causal links through the use of cross-logged panel designs in the field or via experimental manipulation in the laboratory.

References

Abrami, P. C. & Mizener, D. A. Does the attitude similarity of college professors and their students produce "bias" in course evaluations? American Educational Research Journal, 1982, in press.

Abrami, P. C., Perry, R. P., & Leventhal, L. The relationship between student personality characteristics, teacher ratings, and student achievement. Journal of Educational Psychology, 1982, 74, 111-125.

Feldman, K. A. Consistency and variability among college students in rating their teachers and courses: A review and analysis. Research in Higher Education, 1977, 6, 223-274.

Hildebrand, M., & Wilson, R. C. Effective university teaching and its evaluation. Washington, D. C.: U.S. Office of Education, 1970.

Hofman, F. E., & Kremer, L. Attitudes toward higher education and course evaluation. Journal of Educational Psychology, 1980, 72, 610-617.

Leventhal, L., Abrami, P. C., & Perry, R. P. Effects of lecturer quality and student perceptions of lecturer's experience on teacher ratings and student achievement. Journal of Educational Psychology, 1977, 69, 360-374.

Marsh, H. W. Students as evaluators of teaching. In International encyclopedia of education: Research and studies, Pergamon Press, 1983, in press.

Nie, H. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (Eds.). Statistical package for the social sciences (2nd ed.), New York: McGraw-Hill, 1975.

Osgood, C., Suci, G., & Tannenbaum, P. The measurement of meaning. Urbana-Champaign: University of Illinois Press, 1957.

Table 1

Regression of TRF Scores on TOTSIM With

And Without Teacher Effects Removed:

Byrne Attitude Items

CRITERION	SS	MS	F[a]	p <	ΔR^2
		Simple Regression			
TRFA	6.260	6.260	13.666	.001	.044
TRFB	4.210	4.210	19.745	.001	.063
Analytic/Synthetic	2.526	2.526	10.159	.01	.033
Organization/Clarity	6.148	6.148	19.200	.001	.061
Instructor-Group Interaction	6.820	6.820	15.522	.001	.050
Instructor-Individual Student Interaction	4.043	4.043	9.279	.01	.031
Dynamism/Enthusiasm	2.459	2.459	5.891	.05	.020
		Teacher Effects Removed			
TRFA	4.577	4.577	12.471	.001	.032
TRFB	1.769	1.769	11.032	.001	.026
Analytic/Synthetic	1.983	1.983	9.168	.01	.026
Organization/Clarity	2.302	2.302	9.396	.01	.023
Instructor-Group Interaction	1.592	1.592	5.263	.05	.012
Instructor-Individual Student Interaction	1.096	1.096	3.365	ns	-
Dynamism/Enthusiasm	2.006	2.006	7.435	.01	.016

[a] for simple regression, df = 1,294; for stepwise regression df = 1,276.

Table 2

Regression of TRF Scores on TOTSIM With

And Without Teacher Effects Removed:

Hofman & Kremer Attitude Items

CRITERION	SS	MS	F[a]	$p <$	ΔR^2
		Simple Regression			
TRFA	11.607	11.607	23.829	.001	.064
TRFB	8.295	8.295	45.709	.001	.116
Analytic/Synthetic	5.322	5.322	20.279	.001	.055
Organization/Clarity	8.527	8.527	23.278	.001	.063
Instructor-Group Interaction	9.364	9.364	26.719	.001	.072
Instructor-Individual Student Interaction	11.048	11.048	27.126	.001	.073
Dynamism/Enthusiasm	7.784	7.784	20.930	.001	.057
		Teacher Effects Removed			
TRFA	7.018	7.018	18.390	.001	.039
TRFB	3.788	3.788	25.824	.001	.053
Analytic/Synthetic	3.521	3.521	15.322	.001	.037
Organization/Clarity	3.295	3.295	12.723	.001	.024
Instructor-Group Interaction	3.505	3.505	14.872	.001	.027
Instructor-Individual Student Interaction	5.358	5.358	17.520	.001	.035
Dynamism/Enthusiasm	3.431	3.431	12.524	.001	.025

[a] for simple regression, df = 1,347; for stepwise regression, df = 1,326.

Table 3

Stepwise Regression of TRF Scores on Each Set of Component Scores Independent of Teacher Effects: Hofman & Kremer Attitude Items

CRITERION	SS	MS	F[a]	$p <$	ΔR^2
	Own Scores Entered on Last Step				
TRFA	6.202	0.345	0.854	ns	-
TRFB	3.324	0.185	1.300	ns	-
Analytic/Synthetic	6.911	0.384	1.773	.05	.072
Organization/Clarity	6.543	0.364	1.402	ns	-
Instructor-Group Interaction	2.579	0.143	0.613	ns	-
Instructor-Individual Student Interaction	9.803	0.545	2.016	.01	.064
Dynamism/Enthusiasm	6.304	0.350	1.345	ns	-
	Perceived Scores Entered on Last Step				
TRFA	7.631	0.424	1.052	ns	-
TRFB	7.339	0.408	2.870	.001	.103
Analytic/Synthetic	9.767	0.543	2.506	.001	.101
Organization/Clarity	5.869	0.326	1.257	ns	-
Instructor-Group Interaction	8.356	0.464	1.986	.01	.064
Instructor-Individual Student Interaction	16.694	0.927	3.433	.001	.110
Dynamism/Enthusiasm	10.912	0.606	2.328	.01	.080

[a] df = 18,291